

## APPENDIX A

The following is a short, Perl-like pseudocode implementation of an abbreviated intermediary server and several finite state machine implementations. The pseudocode is commented, and is straightforwardly interpreted by anyone skilled in the art of software development.

```
#!/usr/bin/perl -w

10  #
    #
    #

    &start_session_server();
15  die;

20
    #
    #
    #
    sub start_session_server
25  {
        # load configurations
        &load_config_file( "config.txt" );

        # load retrieval keys and FSMs
30  %FSM_HASH = ();
    %ARG_HASH = ();
```

```

&load_FSMs( "FSM_config.txt.pl", \%FSM_HASH, \%ARG_HASH );

# start the server part of Session Server
&start_server( 8080 );
5  }
#
#
#
sub start_server
10  {
    my( $port ) = @_ ;

    &initialize_and_start_server( $port );

15  while( $run == true ){

        # start a new thread
        &listen_and_process_client_request( $url );

20  }

    return;
}

25

30 sub listen_and_process_client_request
    {

```

```

my( $url ) = @_;
my( $rkey ) = &get_retrieval_key_from_url( $url );

my( $fsm ) = &get_corresponding_FSM_name( $rkey, \%FSM_HASH );
5   $_ = $fsm;
    SWITCH: {
        if( /^FSM__session_id$/ ) { &process_FSM__session_id( $rkey; );      last; }
        if( /^FSM__session_id__optimized$/ ) { &process_FSM__session_id__optimized(
10   $rkey; ); last; }
        if( /^FSM__HTML_FORM$/ ) { &process_FSM__HTML_FORM( $rkey; );
        last; }
        # other FSM's can be added here...
    }

15   # exit this thread
    }
    #
    #
    #

20   sub process_FSM__session_id
    {
        my( $rkey ) = @_;

        my( @ARG_ARR ) = split( /\t/, $ARG_HASH{ $rkey } );

25   #
        #   FSM -- first step
        #
        $starturl = shift( @ARG_ARR );
        $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
30   \"$starturl\"`;

```

```

#
#  FSM -- step i
#
5   my( $cnt ) = 1;
    foreach $regexp ( @ARG_ARR ) {
        # $regexp = "<a[^>]+?href=([^\t\r\n]*)[^\>]*>[^\<]*<img[^\>]+alt=\"computers and
peripherals\"";
        if( $doc =~ /$regexp/gsi ) {
10      $nexturl = $1;
          $nexturl =~ s/^[\"\\\"]*//;
          $nexturl =~ s/[\"\\\"]*$//;
          $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
\"$nexturl\"`;
15      } else {
          return "Nexturl at FSM Step $cnt -- cannot be obtained...\n";
        }
        $cnt ++;
    }
20
#
#  FSM -- last step
#
    $base_href = "<BASE HREF=\"\$starturl\">";
25    print $base_href, "\n", $doc, "\n";

    return;
}
#
30 #
#

```

```

sub process_FSM__session_id__optimized
{
    my( $rkey ) = @_;

5    my( @ARG_ARR ) = split( /\t/, $ARG_HASH{ $rkey } );
    #
    #  FSM -- step 0
    #
    $starturl = shift( @ARG_ARR );
10    $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
        \"$starturl\"`;

    #
15    #  FSM -- step 1
    #
    $regex = shift( @ARG_ARR );
    if( $doc =~ /$regex/gsi ){
        $session_ID = $1;
20    }else{
        return "Session ID at FSM Step 2 -- cannot be obtained...\n";
    }

    #
25    #  FSM -- step 2
    #
    $finalurl = shift( @ARG_ARR );
    $regex = shift( @ARG_ARR );
    $finalurl =~ s/$regex/$session_ID/g; # substitute new session ID into the final URL
30

```

```

$doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
\"$finalurl\"";
$base_href = "<BASE HREF=\"$starturl\">";
print $base_href, "\n", $doc, "\n";
5
    return;
}
#
#
10 #
sub process_FSM__HTML_FORM
{
    my( $rkey ) = @_;

15    my( @ARG_ARR ) = split( /\t/, $ARG_HASH{ $rkey } );
    #
    #   FSM -- first step
    #
    $form_url = shift( @ARG_ARR );
20    $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
\"$form_url\"";

    #
25    #   FSM -- step i
    #
    my( $cnt ) = 1;
    foreach $field_value ( @ARG_ARR ){
        ( $field, $value ) = split( /\./, $field_value, 2 );
30        $doc =~ s/$field/$value/g;      # substitute value into the corresponding FORM
field

```

```
        $cnt ++;
    }

    #
5    #  FSM -- last step
    #
    $base_href = "<BASE HREF=\"${form_url}\">";
    print $base_href, "\n", $doc, "\n";

10    return;
}
```

#

# This is the FSM configuration file for the Session Server

#

```

5  sony00001a  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
\t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
\t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>Z1
Series< <a[^>]+?href=([< \t\r\n]*)[>]*>[<]*<img[^>]+alt=\"PCGZ1RAP1KITB\"
sony00001b  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
10 \t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
\t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>Z1
Series< <a[^>]+?href=([< \t\r\n]*)[>]*>[<]*<img[^>]+alt=\"PCG-Z1 VAP2\"
sony00002  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
\t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
15 \t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>V505
Series<
sony00003  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
\t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
\t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>GRT
20 Series<
sony00004  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
\t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
\t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>TR
Series<
25 sony00005  FSM_session_id  http://www.sonystyle.com/  <a[^>]+?href=([<
\t\r\n]*)[>]*>[<]*<img[^>]+alt=\"computers and peripherals <a[^>]+?href=([<
\t\r\n]*)[>]*>[ \t\r\n]*VAIO\&reg\; Notebooks< <a[^>]+?href=([< \t\r\n]*)[>]*>FRV
Series<

```

30



sony\_opt00001a      FSM\_session\_id\_\_optimized http://www.sonystyle.com/  
 <a[^>]+?href=([< \t\r\n]\*)[<]\*>[^<]\*<img[^>]+alt=\"computers and peripherals  
 \;sid=([^\=\\?]+)[\=\\?] http://www.sonystyle.com/is-  
 bin/INTERSHOP.enfinity/eCS/Store/en/-/USD/SY\_DisplayProductInformation-  
 5 Start;sid=\_\_SESSION\_ID\_\_=?CategoryName=cpu\_VAIONotebookComputers\_Z1Series&Pr  
 oductSKU=PCGZ1RAP1KITB&Dept=cpu \_\_SESSION\_ID\_\_

delta\_form00001a      http://www.delta.com/  
 10      (.\*<input[^>]\*?name=\"DEPT\_1\"[<]\*)value=\"\"(.\*)      \$1 value=\"SEA\"\$2  
          (.\*<input[^>]\*?name=\"DEST\_1\"[<]\*)value=\"\"(.\*)      \$1 value=\"SFO\"\$2

```

#!/usr/bin/perl -w

#
#
5  #

$starturl = "http://www.sonystyle.com/";
$doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose \"$starturl\"`;

10
#
#  FSM -- step 1
#
$regexp = "<a[^>]+?href=([> \t\r\n]*)[>]*[<]*<img[^>]+alt=\"computers and
15 peripherals\"";
if( $doc =~ /$regexp/gsi ){
    $nexturl = $1;
    $nexturl =~ s/^[\"\\]*//;
    $nexturl =~ s/[\"\\]*$//;
20  $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
    \"$nexturl\"`;
} else {
    die "Nexturl at FSM Step 1 -- cannot be obtained...\n";
}

25
#
#  FSM -- step 2
#
30 $regexp = "<a[^>]+?href=([> \t\r\n]*)[>]*[ \t\r\n]*VAIO\&reg\; Notebooks<";
if( $doc =~ /$regexp/gsi ){

```

```

$nexturl = $1;
$nexturl =~ s/^[\"\\]*//;
$nexturl =~ s/[\"\\]*$//;
$doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
5 \"$nexturl`;
} else {
    die "Nexturl at FSM Step 2 -- cannot be obtained...\n";
}

10
#
#  FSM -- step 3
#
$regexp = "<a[^>]+?href=([^\t\r\n]*)[^\>]*>Z1 Series<";
15 if( $doc =~ /$regexp/gsi ){
    $nexturl = $1;
    $nexturl =~ s/^[\"\\]*//;
    $nexturl =~ s/[\"\\]*$//;
    $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
20 \"$nexturl`;
} else {
    die "Nexturl at FSM Step 3 -- cannot be obtained...\n";
}

25
#
#  FSM -- step 4
#
$regexp = "<a[^>]+?href=([^\t\r\n]*)[^\>]*>[^\<]*<img[^\>]+alt=\"PCGZ1RAP1KITB\"";
30 if( $doc =~ /$regexp/gsi ){
    $nexturl = $1;

```

```

    $nexturl =~ s/^[\"\\"]*//;
    $nexturl =~ s/[\"\\"]*$//;
    $doc = `wget -O - --load-cookies cookies --save-cookies cookies --non-verbose
\"$nexturl\"`;
5  }else{
    die "Nexturl at FSM Step 4 -- cannot be obtained...\n";
  }

10
  #
  #  FSM -- step 5
  #
  $base_href = "<BASE HREF=\"http://www.sonystyle.com/\">";
15  print $base_href, "\n", $doc, "\n";    # return page to client

  die;

```